

# A large family size is required to detect sexual heterogeneity of the recombination fraction using dominant and codominant DNA-markers\*

L. Gómez-Raya

*Department of Animal Biotechnology. University of Nevada. Reno. Mail Stop 202, Reno, NV 89557. USA*

## Abstract

Maximum likelihood methods for testing sexual heterogeneity of the recombination fraction using dominant and codominant DNA-markers are developed. The methods are tested by Montecarlo computer simulation. Two crosses were investigated 1)  $Ab/aB \times Ab/aB$  and 2)  $Ab/aB \times Ab/\alpha B$  in which A, a, and  $\alpha$  are alleles at a codominant marker and B and b are alleles at a dominant marker. Estimates of the recombination fraction are biased by more than 6 cM if family size is small (50) for the first cross. Also, the empirical rate of type I error under homogeneity ranged between 0.0 and 3.7 when using the tabulated values corresponding at 5%. Even for large family sizes, the rate of type I error was close to 0 when the expected rate was 1 or 5%. Empirical power was very low for the self-fertilization cross requiring large family sizes (2,000) and large differences in the recombination fraction in the two sexes (differences larger than 20 cM). For the second cross ( $Ab/aB \times Ab/\alpha B$ ), both parents were heterozygous but they do not have identical genotypes at the codominant marker. Estimates of the recombination fraction for the two sexes were nearly unbiased in the presence of heterogeneity. The rate of type I error was similar to their expected values for this cross. For a low recombination fraction, empirical power for this cross was higher than 50% for a family size of 500 and a true difference in recombination fraction in the two sexes of approx. 10 cM.

**Additional key words:** dominant markers, heterogeneity, linkage analysis, recombination fraction.

## Resumen

**Se necesita una familia de gran tamaño para estimar la heterogeneidad de la fracción de recombinación entre sexos con marcadores dominantes y co-dominantes**

En este trabajo se desarrollan métodos de máxima verosimilitud para analizar la heterogeneidad de la fracción de recombinación entre sexos utilizando un marcador dominante y otro codominante. Los métodos fueron comprobados utilizando simulaciones de tipo Montecarlo en computador. Se han investigado dos tipos de cruzamiento,  $Ab/aB \times Ab/aB$  y  $Ab/aB \times Ab/\alpha B$ , en los que A, a, y  $\alpha$  son alelos de un marcador codominante, y B y b del marcador dominante. Las estimaciones de la fracción de recombinación resultaron sesgadas por más de 6 cM para familias de tamaño pequeño (50) para el primer cruce. La tasa empírica de error tipo I, asumiendo homogeneidad, fue entre 0,0 y 3,7 cuando se utilizaron los valores tabulados correspondientes al 5%. En familias de tamaño grande, el error tipo I fue próximo a 0 para valores tabulados de 1 ó 5%. La potencia estadística obtenida por simulación en el ordenador fue muy baja para familias con auto-fecundación, requiriéndose familias de tamaño grande (2000) y diferencias grandes en la fracción de recombinación de los dos sexos (diferencias mayores de 20 cM). Para el segundo cruce  $Ab/aB \times Ab/\alpha B$ , los dos progenitores fueron heterocigotos, pero no tuvieron genotipos idénticos para el marcador codominante. Las estimaciones de la fracción de recombinación para los dos sexos fueron mayormente insesgadas en la presencia de heterogeneidad. La tasa de error de tipo I fue similar a su valor esperado para este cruce. Para fracciones de recombinación pequeñas la potencia estadística de este cruce fue de más del 50% para un tamaño familiar de 500 y unas diferencias de 10 cM entre las fracciones de recombinación de los dos sexos.

**Palabras clave adicionales:** análisis de ligamiento, fracción de recombinación, heterogeneidad, marcadores dominantes.

Corresponding author: [lgomezraya@cabnr.unr.edu](mailto:lgomezraya@cabnr.unr.edu)

Received: 01-10-07; Accepted: 18-12-07.

\* This paper is dedicated to Dr. José María Malpica whom the author was very lucky to meet being a student at the INIA in Madrid (Spain). Dr. José María Malpica had huge critical and analytical capacities in solving scientific problems. In addition, he was a model of honesty, principles, and professionalism. Thanks for being an example in my career!

## Introduction<sup>1</sup>

Genetic maps are constructed for most plant and farm animal species. In species in which research resources are large, the genetic markers of choice are usually microsatellites. On the contrary, for species without enough resources for genomics research, such as many plants and aquaculture species, markers widely used consist of Amplified Fragment Length Polymorphism (AFLP) or Random Amplified Polymorphic DNA (RAPD). These markers are essentially dominant, which means that homozygotes for the dominant allele and heterozygotes cannot be distinguished. In addition, genetic maps for those species utilize microsatellites in order to link dominant markers to known locations in the genome. It has been established that linkage analysis using dominant markers are seriously biased when the markers are in repulsion phase and the number of offspring is small (Säll and Nilsson, 1994). It is not known if estimates of the recombination fraction are biased when the linkage analysis includes codominant and dominant loci.

Heterogeneity occurs when there are differences in the recombination fraction for different individuals or for individuals belonging to different sexes (sexual heterogeneity). Evidences of sexual heterogeneity have been reported in several species of plants such as *Pinus radiata* (Moran *et al.*, 1983), maize (Robertson, 1984), *Pinus taeda* (Groover *et al.*, 1995) and species of the genus *Populus* (Yin *et al.*, 2001). Difference in the recombination rate for the two sexes, has also been reported for aquaculture species such as rainbow trout and Atlantic salmon (Sakamoto *et al.*, 2000; Moen *et al.*, 2004).

A general procedure for testing sexual heterogeneity was developed by Wu *et al.* (2002). However, they did not investigate if estimates of the recombination fraction using dominant and codominant markers are biased under sexual heterogeneity. They did not investigate power for alternative family sizes when markers are dominant and codominant in the same linkage analysis.

The objectives of this paper are 1) to develop maximum likelihood equations to test for heterogeneity of the recombination fraction in males and females using a codominant and a dominant marker, 2) to determine by Montecarlo simulation if estimates of the re-

combination fraction in the two sexes are biased when one of the markers is dominant and the other is codominant, 3) to evaluate the magnitude of type I errors empirically in the test of sexual heterogeneity of the recombination fraction, and 4) to compute empirical power of detection of heterogeneity for alternative sizes of a large full-sib family.

## Theory

### Maximum likelihood estimation under sexual heterogeneity using a codominant and a dominant marker

The structure considered is a full-sib family of relatively large size as often found in many plants and fish species. Consider a codominant marker (with alleles A, a and  $\alpha$ ) and a dominant marker (with alleles B and b). Offspring at the dominant marker can be BB, Bb, or bb but individuals BB and Bb cannot be distinguished from each other. Assume that  $c_s$  is the recombination fraction in sires, and  $c_d$  is the recombination fraction in dams. Two types of crosses were considered 1)  $Ab/aB \times Ab/aB$  and 2)  $Ab/aB \times Ab/\alpha B$ . The first corresponds to self-fertilization or the cross between two outbred plants heterozygous at a codominant marker for the same alleles and, therefore, with no informative offspring at the codominant marker. The second allows all offspring to be informative at the codominant marker. These crosses represent all situations where meiosis from both sexes can be used to estimate recombination fractions using one dominant and one codominant marker. For simplicity and given the relatively large size of the full sib family, linkage phases were assumed to be known.

#### 1. Cross $Ab/aB \times Ab/aB$

The gametes contributed by the sire and dam and corresponding frequencies for the genotypes of the offspring are given in Table 1. Using this table, the likelihood equation can be written as:

$$L(\hat{c}_s, \hat{c}_d) = k (\phi_{AAB-})^{n_{AAB-}} (\phi_{AAbb})^{n_{AAbb}} (\phi_{AaB-})^{n_{AaB-}} (\phi_{Aabb})^{n_{Aabb}} (\phi_{aaB-})^{n_{aaB-}} (\phi_{aabb})^{n_{aabb}}$$

<sup>1</sup> Abbreviations used: AFLP (amplified fragment length polymorphism), LRT (likelihood ratio test), MSE (mean square error), RAPD (random amplified polymorphic DNA).

**Table 1.** Gametes from sire and dam and resulting offspring genotypes with their expected frequencies:  $c_s$ : recombination fraction in sires.  $c_d$ : recombination frequency in dams. The cross is  $Ab/aB \times Ab/aB$

Dam	Sire			
	Ab $\frac{1}{2}(1-c_s)$	AB $\frac{1}{2}c_s$	ab $\frac{1}{2}c_s$	aB $\frac{1}{2}(1-c_s)$
Ab $\frac{1}{2}(1-c_d)$	AAbb $\frac{1}{4}(1-c_s)(1-c_d)$	AAbB $\frac{1}{4}c_s(1-c_d)$	aAbb $\frac{1}{4}c_s(1-c_d)$	aAbB $\frac{1}{4}(1-c_s)(1-c_d)$
AB $\frac{1}{2}c_d$	AABb $\frac{1}{4}(1-c_s)c_d$	AABB $\frac{1}{4}c_sc_d$	aABb $\frac{1}{4}c_sc_d$	aABB $\frac{1}{4}(1-c_s)c_d$
ab $\frac{1}{2}c_d$	Aabb $\frac{1}{4}(1-c_s)c_d$	AabB $\frac{1}{4}c_sc_d$	aabb $\frac{1}{4}c_sc_d$	aabB $\frac{1}{4}(1-c_s)c_d$
aB $\frac{1}{2}(1-c_d)$	AaBb $\frac{1}{4}(1-c_s)(1-c_d)$	AaBB $\frac{1}{4}c_s(1-c_d)$	aaBb $\frac{1}{4}c_s(1-c_d)$	aaBB $\frac{1}{4}(1-c_s)(1-c_d)$

where  $k = \text{constant}$ ,  $n_{ij}$  is the gametic count for the offspring with genotype «i» at the codominant marker ( $i = AA, Aa, aa$ ) and genotype «j» at the dominant marker ( $j = B-, bb$ ),  $\phi_{AAB-} = \frac{1}{4}c_sc_d + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{AAbb} = \frac{1}{4}(1-c_d)(1-c_s)$ ,  $\phi_{AaB-} = \frac{1}{2}(1-c_d)(1-c_s) + \frac{1}{2}c_sc_d + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{Aabb} = \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{aaB-} = \frac{1}{4}(1-c_s)(1-c_d) + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ , and  $\phi_{aabb} = \frac{1}{4}c_dc_s$ .

## 2. Cross $Ab/aB \times Ab/aB$

The gametes contributed by the sire and dam and corresponding frequencies for the genotypes of the offspring is given in Table 2. The likelihood equation is:

$$L(\hat{c}_s, \hat{c}_d) = k (\phi_{AAB-})^{n_{AAB-}} (\phi_{AAbb})^{n_{AAbb}} (\phi_{AaB-})^{n_{AaB-}} (\phi_{Aabb})^{n_{Aabb}} (\phi_{aaB-})^{n_{aaB-}} (\phi_{aabb})^{n_{aabb}}$$

where  $k = \text{constant}$ ,  $n_{ij}$  is the gametic count for the offspring with genotype «i» at the codominant marker ( $i = AA, Aa, aa$ ) and genotype «j» at the dominant marker ( $j = B-, bb$ ),  $\phi_{AAB-} = \frac{1}{4}c_sc_d + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{AAbb} = \frac{1}{4}(1-c_d)(1-c_s)$ ,  $\phi_{AaB-} = \frac{1}{2}(1-c_d)(1-c_s) + \frac{1}{2}c_sc_d + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{Aabb} = \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ ,  $\phi_{aaB-} = \frac{1}{4}(1-c_s)(1-c_d) + \frac{1}{4}c_s(1-c_d) + \frac{1}{4}c_d(1-c_s)$ , and  $\phi_{aabb} = \frac{1}{4}c_dc_s$ .

## Testing linkage heterogeneity in the sexes

The testing of heterogeneity of the recombination fraction between sexes can be carried out using a likelihood ratio test (LRT):

$$LRT = -2 \ln \frac{L(\hat{c})}{L(\hat{c}_s, \hat{c}_d)},$$

**Table 2.** Gametes from sire and dam and resulting offspring genotypes with their expected frequencies:  $c_s$ : recombination fraction in sires.  $c_d$ : recombination frequency in dams. The cross is  $Ab/aB \times Ab/aB$

Dam	Sire			
	Ab $\frac{1}{2}(1-c_s)$	AB $\frac{1}{2}c_s$	$\alpha b$ $\frac{1}{2}c_s$	$\alpha B$ $\frac{1}{2}(1-c_s)$
Ab $\frac{1}{2}(1-c_d)$	AAbb $\frac{1}{4}(1-c_s)(1-c_d)$	AAbB $\frac{1}{4}c_s(1-c_d)$	$\alpha Abb$ $\frac{1}{4}c_s(1-c_d)$	$\alpha AbB$ $\frac{1}{4}(1-c_s)(1-c_d)$
AB $\frac{1}{2}c_d$	AABb $\frac{1}{4}(1-c_s)c_d$	AABB $\frac{1}{4}c_sc_d$	$\alpha ABb$ $\frac{1}{4}c_sc_d$	$\alpha ABB$ $\frac{1}{4}(1-c_s)c_d$
ab $\frac{1}{2}c_d$	Aabb $\frac{1}{4}(1-c_s)c_d$	AabB $\frac{1}{4}c_sc_d$	$\alpha abb$ $\frac{1}{4}c_sc_d$	$\alpha abB$ $\frac{1}{4}(1-c_s)c_d$
aB $\frac{1}{2}(1-c_d)$	AaBb $\frac{1}{4}(1-c_s)(1-c_d)$	AaBB $\frac{1}{4}c_s(1-c_d)$	$\alpha aBb$ $\frac{1}{4}c_s(1-c_d)$	$\alpha aBB$ $\frac{1}{4}(1-c_s)(1-c_d)$

where  $L(\hat{c}_s, \hat{c}_d)$  is the maximum likelihood under heterogeneity and  $L(\hat{c})$  is the maximum likelihood under homogeneity, i.e, when  $c_s = c_d$ . LRT is distributed as a  $\chi^2$  with one degree of freedom.

## Monte Carlo simulation

A full-sib family with varying offspring size (50, 100, 500, 2000) was simulated in order to investigate how progeny size might affect bias, type I error, and power of detection of heterogeneity of the recombination fraction in the two sexes. It was assumed that linkage phase was known. The probability of transmission of either allele at each marker from sire and dam to offspring was assumed to be  $1/2$ .

Offspring from the sire was simulated using random drawings from a uniform distribution. For the cross  $Ab/aB \times Ab/aB$  the generation of the gametic counts for sires was as follows. If a drawing was in the interval between 0 and  $1/2 (1-c_s)$ , then the gametic count for  $Ab$  was increased by one. If the drawing was between  $1/2 (1-c_s)$ , and  $1/2$  then the gametic count for genotype  $AB$  was increased by one. If a drawing was in the interval between  $1/2$  and  $1/2 (1+c_s)$ , then the gametic count for  $ab$  was increased by one. If the drawing was between  $1/2 (1+c_s)$ , and 1 then the gametic count for genotype  $aB$  was increased by one. The gametes from the dams were generated following the same procedure. The genotypes in the offspring were constructed using sire and dam contributions. The offspring genotypes from the cross  $Ab/aB \times Ab/aB$  were generated following the same principles. Progeny genotypic counts were used to estimate recombination fraction using the grid search method. Each simulation set was replicated 1,000 times. Simulated recombination fractions for sires and dams were 0.05, 0.15, 0.25 and 0.35 and all combinations of them. LRT for testing heterogeneity of the recombination fraction in the sexes was computed for each replicate.

Bias for the estimates of the recombination fraction in sex  $j$  was computed as:

$$Bias(\hat{c}_j) = \frac{\sum_{i=1}^{rep} \hat{c}_{ji}}{rep} - c_j,$$

where  $\hat{c}_{ji}$  is the estimate of the recombination fraction for the  $i$ -th replicate and the  $j$ -th sex ( $j = 1, 2$ ),  $rep$  is the number of replicates in the simulation and  $c_j$  is the true (simulated) recombination fraction in sex  $j$ . Variance of the estimates of the recombination fraction for sex  $j$  was computed as:

$$V(\hat{c}_j) = \frac{\sum_{i=1}^{rep} \hat{c}_{ji}^2}{rep} - \left( \frac{\sum_{i=1}^{rep} \hat{c}_{ji}}{rep} \right)^2.$$

Mean square error (MSE) of the estimates of the recombination fraction for sex  $j$  was computed as:

$$MSE(\hat{c}_j) = \sum_{i=1}^{rep} (\hat{c}_{ji} - c_j)^2 / rep.$$

Empirical rate of type I error was computed in simulations sets where simulated (true) recombination fractions were identical in males and females by computing the number of replicates exceeding the thresholds at  $P < 0.01$  and  $P < 0.05$  corresponding to a  $\chi^2$  with 1 degree of freedom and by dividing those values by the total number of replicates. Empirical power was computed in the simulations in which simulated recombination fractions for the male and the female were different, and by computing the number of replicates that were significant after using the tabulated values divided by the total number of replicates.

## Results

First, recombination fractions using the full model (sexual heterogeneity) under the null hypothesis (homogeneity) were estimated. Estimates were generally biased in the cross  $Ab/aB \times Ab/aB$  and unbiased for the cross  $Ab/aB \times Ab/aB$ . For example, bias in the cross  $Ab/aB \times Ab/aB$  for a family size of 50 varied from 2 to 12 cM.

Second, estimates of the recombination fractions in the two sexes using the full model and the alternative hypothesis (heterogeneity) were carried out. The estimates of recombination fractions in sires and dams, and their variances, together to the mean square error are presented in Table 3 for the cross  $Ab/aB \times Ab/aB$  and in Table 4 for the cross  $Ab/aB \times Ab/aB$ . If the family size is relatively small (50 or 250 offspring) then the estimates of the recombination fraction in the two sexes for the  $Ab/aB \times Ab/aB$  cross are biased. The bias can also be noticed by comparing the values of the variances of the estimates of the recombination fraction with their corresponding mean square error. For small family sizes, the bias can be more than 0.06 units of recombination fraction (approx. 6 cM) and tend to be negative. Therefore, the loss of information when both parents are heterozygous for the same alleles

**Table 3.** Simulation results from the cross between Ab/aB  $\times$  Ab/aB. Bias in estimating recombination fraction in sires [Bias( $\hat{c}_s$ )], and dams [Bias( $\hat{c}_d$ )], variance of the estimates of the recombination fractions for sires [V( $\hat{c}_s$ )], and dams [V( $\hat{c}_d$ )], and mean square error of the estimates for sires [mse( $\hat{c}_s$ )], and dams [MSE( $\hat{c}_d$ )].  $c_s$  = simulated recombination fraction in sires.  $c_d$  = simulated recombination fraction in dams. 100 multiply all values in the table but the size. Size = number of full-sib offspring in the family

Size	$c_s$	Bias ( $\hat{c}_s$ )	V( $\hat{c}_s$ )	MSE( $\hat{c}_s$ )	$c_d$	Bias( $\hat{c}_d$ )	V( $\hat{c}_d$ )	MSE( $\hat{c}_d$ )
50	5.00	-3.43	17.28	29.05	15.00	2.82	64.69	72.65
50	5.00	-2.09	46.47	50.84	25.00	1.77	86.66	89.77
50	5.00	-0.8	73.98	74.61	35.00	-0.2	100.95	100.99
50	15.00	-6.8	120.99	167.28	25.00	6.04	108.40	144.83
50	15.00	-3.22	167.31	177.68	35.00	1.8	109.68	112.92
50	25.00	-6.59	211.23	254.69	35.00	4.41	100.26	119.67
250	5.00	-1.25	24.19	25.77	15.00	0.92	24.49	25.34
250	5.00	0.56	36.02	36.33	25.00	0	24.29	39.15
250	5.00	0.67	34.46	34.91	35.00	-0.72	41.83	42.35
250	15.00	-1.88	55.93	59.46	25.00	1.54	48.45	50.81
250	15.00	0.44	73.95	74.14	35.00	-0.75	64.43	65.00
250	25.00	-2.34	73.99	79.45	35.00	1.9	61.39	65.00
500	5.00	-0.21	18.62	18.67	15.00	0.22	19.48	19.53
500	5.00	0.54	22.30	22.59	25.00	-0.62	25.20	25.59
500	5.00	0.19	15.26	15.30	35.00	-0.27	20.85	20.92
500	15.00	-1.17	38.87	40.24	25.00	0.87	34.93	35.69
500	15.00	0.5	46.58	46.84	35.00	-0.74	43.24	43.79
500	25.00	-1.62	48.94	51.57	35.00	1.23	45.52	44.03
2000	5.00	0.3	10.04	10.13	15.00	-0.31	9.89	9.99
2000	5.00	0.17	6.64	6.66	25.00	-0.14	7.11	7.13
2000	5.00	0.02	3.46	3.46	35.00	0.06	4.57	4.58
2000	15.00	-0.13	20.54	20.56	25.00	0.05	18.78	18.78
2000	15.00	0.59	19.90	20.26	35.00	-0.56	18.29	18.60
2000	25.00	-0.21	25.81	25.86	35.00	0.19	23.45	23.49

at the codominant marker affects the estimation of the recombination fraction. For large family sizes (500 or 2000), estimates of the recombination fractions are nearly unbiased since the values of the average of the estimates of recombination fraction and the simulated are similar. Also, there is a good agreement between the variance of estimates of the recombination fraction of the two sexes with their respective mean square error. If the codominant marker is fully informative (cross Ab/ $\alpha$ B  $\times$  Ab/aB) then the estimates of the recombination fractions in the two sexes are nearly unbiased (Table 4). Bias is 0.0082 or less for all situations simulated. There is a reduction in the bias as family size increases. In addition, variances of the estimates of the recombination fractions were very similar to their corresponding mean square error.

The next step was to evaluate if the maximum likelihood method using tabulated values is appropriate for testing heterogeneity of the recombination fraction between sexes. The empirical percentage of significant results at  $P < 0.01$  and  $P < 0.05$  when simulating the

null hypothesis (homogeneity) in a variety of scenarios is depicted in Table 5. That is, the empirical type I error was computed for each simulation set and compared to their expected 1 or 5%. The performance of the cross Ab/aB  $\times$  Ab/aB to detect heterogeneity when in reality does not exist is clearly low for any sample size. For example, for an expected rate of type I error of 5%, the empirical values in the simulation ranged between 0.0 and 3.7. It seems that the rate of type I error is much lower than expected when the recombination fraction is below 0.05. The estimates of the recombination fractions under homogeneity were in good agreement with those simulated in the cross Ab/ $\alpha$ B  $\times$  Ab/aB. The rate of type I error was generally in good agreement with their expected values.

A next relevant question is to know which family size is needed to detect heterogeneity under true heterogeneity. The empirical power computed as the percentage of significant results at  $P < 0.05$  and  $P < 0.01$  for both types of crosses are presented in Table 6. Power to detect heterogeneity in the cross is very low



**Table 4.** Simulation results from the cross  $Ab/\alpha B \times Ab/aB$ , where  $\alpha$  is an allele different from A and a. Bias in estimating recombination fraction in sires [ $Bias(\hat{c}_s)$ ], and dams [ $Bias(\hat{c}_d)$ ], variance of the estimates of the recombination fractions for sires [ $V(\hat{c}_s)$ ], and dams [ $V(\hat{c}_d)$ ], and mean square error of the estimates for sires [ $mse(\hat{c}_s)$ ], and dams [ $MSE(\hat{c}_d)$ ].  $c_s$  = simulated recombination fraction in sires.  $c_d$  = simulated recombination fraction in dams. 100 multiply all values in the table but the size. Size = number of full-sib offspring in the family

Size	$c_s$	Bias ( $\hat{c}_s$ )	V( $\hat{c}_s$ )	MSE( $\hat{c}_s$ )	$c_d$	Bias( $\hat{c}_d$ )	V( $\hat{c}_d$ )	MSE( $\hat{c}_d$ )
50	5.00	0.31	36.37	36.46	15.00	-0.82	69.95	70.62
50	5.00	0.06	36.82	36.82	25.00	-0.19	86.08	86.12
50	5.00	0.08	35.99	35.99	35.00	-0.76	89.46	90.03
50	15.00	0.12	90.64	90.66	25.00	-0.02	102.37	102.37
50	15.00	0.26	88.82	88.89	35.00	-0.63	102.94	103.34
50	25.00	0.43	112.05	112.23	35.00	-0.79	110.25	110.88
250	5.00	-0.08	6.94	6.94	15.00	-0.14	11.85	11.87
250	5.00	0.01	7.64	7.64	25.00	-0.12	16.11	16.13
250	5.00	-0.06	7.39	7.40	35.00	-0.03	19.64	19.64
250	15.00	0.02	16.82	16.82	25.00	-0.09	18.45	18.46
250	15.00	-0.12	17.66	17.68	35.00	-0.03	21.20	21.21
250	25.00	-0.1	24.57	24.58	35.00	0.1	25.61	25.62
500	5.00	-0.05	3.27	3.27	15.00	0.12	5.96	5.98
500	5.00	-0.06	3.30	3.30	25.00	0	8.53	8.53
500	5.00	-0.08	3.40	3.41	35.00	-0.01	10.20	10.20
500	15.00	-0.06	8.01	8.01	25.00	-0.03	9.91	9.91
500	15.00	-0.08	8.12	8.13	35.00	-0.05	11.84	11.84
500	25.00	-0.09	11.17	11.17	35.00	-0.06	12.65	12.66
2000	5.00	-0.02	0.81	0.81	15.00	0.04	1.50	1.50
2000	5.00	-0.03	0.84	0.84	25.00	0.05	2.06	2.06
2000	5.00	-0.02	0.86	0.86	35.00	0.08	2.29	2.30
2000	15.00	-0.02	2.16	2.16	25.00	0.03	2.34	2.34
2000	15.00	-0.02	2.26	2.26	35.00	0.06	2.66	2.66
2000	25.00	0.01	3.05	3.05	35.00	0.07	2.87	2.88

for all cases considered except when the size of the offspring is large and the differences between sexes are also large. For example, for 2000 offspring and recombination fractions of 0.05 and 0.35, the power is 88.5 and 96.4 for significance levels at  $P < 0.01$  and  $P < 0.05$ , respectively. This amount of offspring is feasible in many plants and fish species. However, it may not be practical the use of such large family size given the cost and effort required. On the other hand, power to detect heterogeneity at a significance level of  $P < 0.05$  is higher than 50% for the cross  $Ab/\alpha B \times Ab/aB$  for differences in the sex recombination rate of 10 units (approx. 10 cM) and family sizes of 500 except when recombination rates in the two sexes were high (over 0.25).

## Discussion

Heterogeneity of the recombination fraction between sexes using linkage information on dominant and

codominant markers is investigated in this paper. The first step was to construct maximum likelihood equations allowing testing heterogeneity of the recombination fraction in the two sexes. The methods were implemented in software and tested by computer simulation. In a second step, bias in estimating the recombination fraction was investigated. It can be concluded from the results of the simulation experiments using crosses  $Ab/aB \times Ab/aB$  that 1) sexual heterogeneity can be detected when in reality there is not heterogeneity, and 2) estimation of the recombination fraction is biased for small progeny groups. Bias in estimating recombination fraction has already been reported when the two markers are dominant and in a repulsion phase (Säll and Nilsson, 1994) or when neglecting non-informative offspring using half-sib offspring without genotype information from the dams (Gómez-Raya, 2001). The bias reported in this paper is likely due to limited family size since it decreases as family size increases in the simulation. This is a general problem of the maximum

**Table 5.** Empirical rate of Type I error when testing herogeneity between sexes at significance level of 0.05 ( $P < 0.05$ ) and at significance level of 0.01 ( $P < 0.01$ ) for true homogeneity of the recombination fraction between sexes and for varying sizes in the offspring in crosses  $Ab/aB \times Ab/aB$  and  $Ab/\alpha B \times Ab/aB$ . c: simulated recombination fraction. All values in the table but the size are multiplied by 100

c	Size	$Ab/aB \times Ab/aB$		$Ab/\alpha B \times Ab/aB$	
		$P < 0.01$	$P < 0.05$	$P < 0.01$	$P < 0.05$
5.00	50	0.0	0.0	0.0	1.6
	250	0.0	0.0	1.5	6.8
	500	0.0	0.0	0.9	5.7
	2000	0.0	0.5	1.1	5.3
15.00	50	0.0	0.1	1.3	8.3
	250	0.0	2.4	1.0	4.6
	500	0.4	3.7	1.0	5.0
	2000	1.5	3.7	1.0	5.8
25.00	50	0.0	0.3	1.3	6.6
	250	1.2	3.6	0.8	4.3
	500	0.6	2.5	1.0	4.7
	2000	1.3	3.0	1.1	4.5
35.00	50	0.0	0.6	1.4	5.1
	250	0.2	1.4	0.8	4.5
	500	0.1	2.3	0.9	4.4
	2000	0.4	2.5	0.9	5.0

likelihood estimators that are only asymptotically unbiased for large sample sizes. Linkage reports might provide estimates of recombination fractions (and genetic distances) with different bias depending on family size of each study. In addition, the required family size for unbiased estimation of the recombination fraction is prohibited for economical reasons. It is more cost efficient the typing of more codominant markers in the chromosomal area such as microsatellites. Bias for the other cross ( $Ab/\alpha B \times Ab/aB$ ) was much smaller even for small family sizes.

A third step was to evaluate the magnitude of the empirical type I error for the two crosses investigated. The computations in the simulation made use of thresholds values tabulated from a  $\chi^2$  distribution. This represents the most common situation in which the researcher performs the test using tabulated values. More advance computational methods can use permutations of the same data to generate the distribution of the parameter under the null hypothesis. The underestimation of the empirical thresholds values in the first type of cross would support the need of such methods in those situations even for relative large family sizes. However, a general good agreement was found for the

**Table 6.** Empirical power for detection of heterogeneity between sexes at significance level of 0.05 ( $P < 0.05$ ) and at significance level of 0.01 ( $P < 0.01$ ) for different true recombination fraction in sire ( $c_s$ ) and dams ( $c_d$ ) and for varying sizes in the offspring

$c_s$	$c_d$	Size	$Ab/aB \times Ab/aB$		$Ab/\alpha B \times Ab/aB$	
			$P < 0.01$	$P < 0.05$	$P < 0.01$	$P < 0.05$
0.05	0.15	50	0.0	0.1	1.9	15.0
		250	0.0	0.0	26.1	50.7
		500	0.0	2.5	60.7	83.4
		2000	3.1	11.6	99.9	100.0
0.05	0.25	50	0.0	0.0	13.3	38.2
		250	0.2	9.1	83.2	93.7
		500	5.9	20.1	99.1	99.7
		2000	29.1	52.8	100.0	100.0
0.05	0.35	50	0.0	1.3	33.4	60.8
		250	9.0	32.6	99.1	99.9
		500	23.1	47.3	100.	100.
		2000	88.5	96.4	100.0	100.0
0.15	0.25	50	0.0	0.4	3.7	12.5
		250	0.7	6.2	14.8	33.7
		500	1.4	4.4	33.0	58.1
		2000	1.8	7.4	95.70	99.30
0.15	0.35	50	0.0	0.8	10.7	24.1
		250	2.5	9.4	65.2	84.3
		500	2.8	9.8	93.9	98.4
		2000	12.0	27.2	100.0	100.0
0.25	0.35	50	0.0	0.9	3.20	8.90
		250	0.3	3.9	11.70	29.00
		500	0.4	3.8	24.50	47.20
		2000	1.4	5.5	90.50	97.40

cross  $Ab/\alpha B \times Ab/aB$ , which would support the use of tabulated values.

On the other hand, linkage phase was assumed to be known in this study. This may not be the general situation found in practice. However, the required large family sizes used for testing heterogeneity would give much weigh determining the maximum likelihood of the most likely phase compared to the maximum likelihood of the less likely phase. Consequently, the impact of this assumption on the results must be rather small.

The general conclusions of this paper is that linkage heterogeneity in the two sexes can be reliably detected using family sizes of at least 500 offspring as long as the genotypes at the codominant marker are not identical in the two parents. For smaller family sizes and other genotype configurations power for detection is only high when the difference in the recombination fraction in the two sexes is rather large (more than 20 cM). The-

refore, true sexual heterogeneity for the recombination fraction might be undetected in experiments using small family sizes.

## References

- GÓMEZ-RAYA L., 2001. Biased estimation of the recombination fraction using half-sib families and informative offspring. *Genetics* 157, 1357-1367.
- GROOVER A.T., WILLIAMS C.G., DEVEY M.E., LEE J.M., NEALE D.B., 1995. Sex-related differences in meiotic recombination frequency in *Pinus taeda*. *J Hered* 86, 157-158.
- MOEN T., HOYHEIM B., MUNCK H., GÓMEZ-RAYA L., 2004. A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Anim Genet* 35(2), 81-92.
- MORAN G.F., BELL J.C., HILLIKER A.J., 1983. Greater meiotic recombination in male vs female gametes in *Pinus radiata*. *J Hered* 74,62.
- ROBERTSON D.S., 1984. Different frequency in the recovery of crossover products from the male and female gametes of plants hypoploid for B-A translocations in maize. *Genetics* 107, 117-130.
- SAKAMOTO T., DANZMANN R.G., GHARBI K., HOWARD P., OZAKI A., KHOO S.K., WORAM R.A., OKAMOTO N., FERGUSON M.M., HOLM L.E., GUYOMARD R., HOYHEIM B., 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155, 1331-1345.
- SÄLL T., NILSSON N.O., 1994. The robustness of recombination frequency estimates in intercrosses with dominant markers. *Genetics* 137, 589-596.
- WU R., CHANG-XING MA, WU S.S., ZENG Z.B., 2002. Linkage mapping of sex-specific differences. *Genet Res Camb* 79, 85-96.
- YIN T.M., ZHANG X.Y., HUANG M.R., WANG M.X., ZHUGE Q., ZHU L.H., ZENG Z.B., WU R.L., 2001. Molecular linkage maps of the *Populus* genome. *Genome* 45, 541-555.